

Why does Dupehunter Professional use XML?

Last revised September 2005, version no. 4.0.0.1720
©Copyright 1999-2005 Carsten Heidtke Software

An **XML export** is very fast, and quite useful for exporting large result lists of 10 million files. Using the XML format, a large number of values can be detected and managed. Weak spots can be examined and exported to external databases.

Revisions:

Dupehunter Professional offers 2 different versions of support for XML export. This means two different ways of implementing XML in Dupehunter. For compatibility reasons, we have retained Revision 1 for companies currently using this format.

With the addition of Revision 2, we have tried to optimize the output with additional tags and values. Revision 2 is recommended. New features include the **Unique IDentifier** tag **<UID>**. This can be used to identify documents and to order multiple export files by creation time. The UID contains the date and time of the export, beginning with the year and ending with seconds. All files exported using Revision 2 are named using the pattern **<UID>.xml** and stored in the export directory. Results exported using Revision 1 are given the name **results.xml**.

Priorities:

In corporate or institutional settings, large numbers of files could contain many duplicates. If several thousand files are found in a single search, it is possible to lose sight of important files, such as system files, that may appear on the list.

Dupehunter Professional supports priorities for each file extension that you have defined. For example, you can add the **.exe** extension and assign a priority value of 100, then add the **.bmp** extension with a priority value of 5. When Dupehunter exports the search results, matched extensions will include the **<PRIORITY>** tag and the priority value. All other files with no special priority definition have no **<PRIORITY>** tag.

You can now search and examine the exported result lists with other third-party products.

Please note:

Always include the dot for each file extension entered: e.g. **.bmp** and not **bmp**.

Because an analysis using priorities requires both additional time and disk space, this option can be enabled or disabled by changing the **Report Priority** option.

Compression and creation of self-extracting archives:

XML files are larger than binary file formats, but they can be compressed with a **high compression ratio**. The integrated SQX archiver in Dupehunter Professional can automatically compress the exported files, and after performing an integrity check using the CRC-32 algorithm, can automatically delete the source file. The compression works with several million files and huge export files. A source file of 250 megabytes can be compressed into a small 3-megabyte archive. If you like, you may also create self-extracting archives.

Additional types of exports, detailing file operations such as deletion, movement or compression, are currently under development, and will be released as soon as possible.